Department of Biostatistics and Medical Informatics Seminar

Rafael Ferreira, Ph.D.

Data Scientist
Data Science Institute
University of Wisconsin–Madison
https://dsi.wisc.edu/staff/ferreira-rafael/

Friday, March 27, 2026
12:00-1:00 pm
Morgridge Hall Seminar Room 2516

Zoom Meeting: 966 3372 9112
https://uwmadison.zoom.us/j/96633729112?pwd=tHFc9i1dAAgmXe05uWtw8wXBIQZxGB.1
Passcode: 621125

Multi-Modal Machine Learning for the Early Detection of Subclinical Ketosis in Dairy Cattle

Abstract: Subclinical ketosis (SCK) is a prevalent metabolic disorder in dairy cows during the transition period, leading to significant economic losses and compromised animal welfare. Early and automated detection of SCK can enable timely preventive interventions, yet it remains challenging due to the absence of visible clinical signs and the multifactorial nature of the disease. This presentation covers two complementary projects aimed at developing multi-modal machine learning systems for the early detection of SCK using exclusively prepartum data to predict postpartum risk.

The first project explores the integration of computer vision, wearable sensor data, and large language models (LLMs) for SCK prediction. Depth videos from 92 dairy cows were collected weekly during the last three weeks before calving, and body shape features were extracted using three approaches: a convolutional neural network (CNN) trained for body condition score (BCS) prediction, depth value sampling between automatically detected anatomical keypoints, and a CNN coupled with a recurrent neural network (CNN-RNN) for temporal BCS feature extraction. These image-derived features were combined with prepartum behavioral data from wearable sensors, cow history information, and text embeddings extracted from farm management software notes using both generic and fine-tuned LLMs. Random Forest models integrating image and tabular data achieved F_1 scores of 0.706, outperforming models using either source alone. Incorporating text embeddings from a fine-tuned LLM further improved performance when coupled with sensor data (0.681 vs 0.655), demonstrating the potential of natural language processing techniques for extracting valuable information from unstructured farm records.

The second project extends this work by incorporating genomic data and proposing a cloud computing-based framework for automated, scalable data processing and integration. Genotypic information comprising 78,964 single-nucleotide polymorphisms per cow was reduced to 128 dimensions via UMAP and combined with imaging and sensor features. A modular pipeline was developed for image processing, including body segmentation, frame quality assessment, and animal identification. Three data fusion strategies (early fusion, late fusion, and cooperative learning) were evaluated, with late fusion achieving the highest F_1 scores (up to 0.750) for binary SCK classification. The cloud-based architecture features independent, reusable services that can be customized for diverse precision livestock farming applications. Together, these projects demonstrate that integrating phenomic and genomic data through modern machine learning and data fusion techniques can substantially improve early SCK detection, supporting the implementation of preventive health management practices in dairy farms.

Bio: Rafael Ferreira is an Electrical and Computer Engineer and Data Scientist at the Data Science Institute at the University of Wisconsin-Madison. He earned his master's in computer science and his PhD in dairy science from the University of Wisconsin-Madison, specializing in developing machine learning and computer vision systems to improve dairy farm management and phenotyping. His research focused on animal identification and early disease detection in dairy cows by integrating data from multiple sensors and incorporating computer vision and natural language processing techniques into a multi-modal machine learning pipeline.

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN–MADISON